

DATABASE

TRENDS AND APPLICATIONS

Solutions for the Information Project Team • www.dbta.com

Volume 18, Number 11 • November 2004

QUICK LOOKS



JAMES F. KOOPMANN

WebQL Brings Structure and Order to Unstructured Environments

QL2 Software and WebQL have the ability to completely automate the scheduling, crawling, extraction, and formatting of unstructured and hidden information through robust harvesting techniques. The largest repository of unstructured data is the Web. Web sites are complex, loaded with dynamic pages, serve up many document types, execute scripts when visited, and require forms to be filled out. WebQL's unique ability to handle dynamically generated Web pages, forms, and scripts facilitates exploration and discovery of deeply hidden Web pages, forms, and documents that hold the vast majority of valuable information. After extraction, WebQL will transform information into a format that is suited for the customer's viewing or integration preference. During the harvesting stage, WebQL provides visual monitoring of the extraction process, shows transfer rates, provides complete logging and notification, and user specified or auto-detect throttling of internal parallel processing to compensate for Web site performance. While working with WebQL, some knowledge of Web

sites is needed, especially when providing information for forms or when tricky extraction techniques must be employed. Simply pull the Web page source, investigate the form, and then re-submit the original query with a specified set of values to be used.

WebQL aids developers to work through the technical complexity of mining through intricate Web sites and documents. In particular, the recently released WebQL version 2.3 is the only software tool that has unlocked the complexities of PDF data extraction with 100 percent accuracy. Developers become quickly productive because WebQL provides for and follows many of the industry standards. There are no unusual coding or content standards that have to be re-learned. If developers understand ANSI SQL, some HTML and XML, and maybe a programming language such as Java or C++, they will immediately begin to feel right at home with WebQL's IDE and be able to produce something immediately.

Customers purchasing the WebQL product for internal use account for 40

percent of QL2's business. While the other 60 percent of QL2's customers purchase custom QL2 Solutions to solve specific unstructured data challenges. QL2 Solutions can be run in-house or on the QL2 Client Center as an ASP delivered applications. A vendor that actually uses its product to develop applications for over half of their customer base is a good indicator that the product works and is well supported. Future QL2's efforts are now concentrating on how they can leverage and complement their "holy grail" of data extraction with tools to analyze, organize, and present the extracted information. This permits QL2 to offer a holistic solution and answers the questions of what to do after WebQL has become the foundation for a company's unstructured data extraction. QL2 Software may be reached at www.ql2.com.

James F. Koopmann has worked with a variety of database-centric software and tools vendors as strategist, architect, and performance expert. He is founder and president of Pine Horse, Inc. Contact him at jkoopmann@pinehorse.com.